Equations of Motion
from a
Data Series

James P. Crutchfield and Bruce S. McNamara
Physics Department
University of California
Berkeley, California 94720 USA

Abstract. Temporal pattern learning, control and prediction, and chaotic data analysis share a common problem: deducing optimal equations of motion from observations of time-dependent behavior. Each desires to obtain models of the physical world from limited information. We describe a method to reconstruct the deterministic portion of the equations of motion directly from a data series. These equations of motion represent a vast reduction of a chaotic data set's observed complexity to a compact, algorithmic specification. This approach employs an informational measure of model optimality to guide searching through the space of dynamical systems. As corollary results, we indicate how to estimate the minimum embedding dimension, extrinsic noise level, metric entropy, and Lyapunov spectrum. Numerical and experimental applications demonstrate the method's feasibility and limitations. Extensions to estimating parametrized families of dynamical systems from bifurcation data and to spatial pattern evolution are presented. Applications to predicting chaotic data and the design of forecasting, learning, and control systems, are discussed.

1. Introduction

When refining a model of a physical process a scientist focuses on the agreement of theoretically predicted and experimentally observed behavior. If these agree in some accepted sense, then the model is "correct" within that context. Lorenz[1, 2] pointed out the fundamental limitations to which this scientific procedure is subject, when the underlying physical process is chaotic. In considering exact theoretical prediction, there

---

Correspondence regarding this work should be directed to the first author. The electronic mail address for this is chaos at gojira.berkeley.edu.

BSMs permanent address: Physics Board of Studies, University of California, Santa Cruz, California 95064.

is an irreducible long-term error in the prediction of system's state that is on the order of the chaotic attractor's size in state space.[3] Even a "correct" model cannot *exactly* reproduce observed chaotic behavior, and one must turn to geometric[4] or statistical[5] criteria of validity. Such qualitative information about the geometry and the asymptotic distribution of states on the attractor is contained, in fact, in a single time series.[4, 6] At present, a model of a chaotic process is considered "correct" when the geometry, dimension, and entropy, of its attractor agree with those of the attractor reconstructed from the observed data.

Here we consider the inverse problem to verifying theoretical models: how can we obtain the equations of motion directly from measurements? To do this we shall extend the notion of qualitative information contained in a sequence of observations to consider directly the underlying dynamics. We will show that using this information one can deduce the effective equations of motion. The latter summarize up to an *a priori* specified level of correctness, or accuracy, the deterministic portion of the observed behavior. The observed behavior on short time scales unaccounted for by the reconstructed equations will be considered extrinsic noise.[7]

This model *ansatz* is tantamount to assuming that the observations have been produced by some arbitrary dynamical system in the presence of fluctuations. Unless explicitly noted, we restrict our discussion to the case that the dynamical system is finite-dimensional: the state is specified by the point $\vec{x} = (x_0, x_1, \cdots, x_{m_{bed}-1}) \in \mathbf{M}$, where $\mathbf{M}$ is m_{bed}-dimensional manifold, the *state space.* The state evolves according to the *dynamic* $\vec{F}(\vec{x}) = (f_0(\vec{x}), f_1(\vec{x}), \cdots, f_{m_{bed}-1}(\vec{x}))$. If the behavior varies continuously in time then the system evolves according to a stochastic differential equation

$$\dot{\vec{x}}\,(t) = \vec{F}(\vec{x}\,(t)) + \xi(t) \quad . \tag{1.1}$$

In the case of discrete-time, the evolution is specified by a stochastic difference equation

$$\vec{x}_{n+1} = \vec{F}(\vec{x}_n) + \vec{\xi}_n \quad , \tag{1.2}$$

where n denotes time step. Further assumptions appear in these equations: the effect of fluctuations is additive,.nr 0C 0 independent of $\vec{x}$. We take $\vec{\xi}\,(t)$, for example, to be zero-mean Gaussian distributed δ-correlated noise with amplitudes $\vec{\sigma}_{ext}$ along each coordinate: $< \xi(t)\xi(t-\tau) > = \delta(\tau)\,\vec{\sigma}_{ext}$. A *model M* then consists of the pair $(\vec{F}, \vec{\sigma}_{ext})$ of

---

$\tau << \dfrac{I_o}{h_\mu}$, *where* h_μ *is the metric entropy and* $I_o \simeq -\,log_2\epsilon$ *is the unbiased average information obtained from a measurement of resolution* ϵ.

relevant dynamics, with concomitant increase in the model's complexity, the extrinsic noise may be revealed to be of deterministic origin and therefore incorporated into the estimated equations of motion.

fluctuation coupling are appropriate, experience has shown that the additive form is adequate for most modeling purposes. The analysis presented here may be carried through for these other models with greater theoretical and computational difficulty.

estimated dynamic and extrinsic noise level vector.

To aid in the geometric interpretation of a model $M = (\vec{F}, \vec{\sigma}_{ext})$ we introduce slightly more general terminology. If the space of all dynamical systems on some manifold $\mathbf{M}$ is denoted $\mathbf{D}(\mathbf{M})$: then $\vec{F} \in \mathbf{D}(\mathbf{M})$; that is, the model dynamic is a point in the space of (deterministic) dynamical systems. In practice we must consider a wider class of dynamical systems as we initially do not know the dimension of $\mathbf{M}$, and so $\mathbf{D}$ shall denote the space of all dynamical systems, and $\vec{F}$ is to be considered a point in this larger space also. The model M with its stochastic component is then an "ellipsoid" centered on $\vec{F} \in \mathbf{D}$. The set of dynamical systems in this ball are *noise-equivalent,* as they describe essentially the same time-averaged dynamics in the presence of a given level $\vec{\sigma}_{ext}$ of fluctuations. Said another way, M is an ensemble of realizations of a stochastic dynamical system. In practice we introduce reconstruction basis coordinates for $\mathbf{M}$ and a function basis $\vec{\phi}(\vec{x}) = (\phi_0(\vec{x}), \cdots, \phi_{K-1})$ for approximating the dynamic $\vec{F}$ to "order" K: $\vec{F}(\vec{x}) = A\,\vec{\phi}(\vec{x})$, where $a_{ik} = (A)_{ik}$ is the coefficient of the k.lf 364 ϕ_k for f_i. These choices result in the space of deterministic models $\mathbf{D}_M$ which is the Km_{bed}-dimensional space of approximations to the infinite-dimensional $\mathbf{D}$. Note that we do not include the m_{bed} components of $\vec{\sigma}_{ext}$ in $\mathbf{D}_M$ and that $\vec{F}$ is linear in the parameters $\{a_{ik}\}$.

Properly considered, the problem of deducing the deterministic portion of a data series is a subset of the general problem of pattern recognition: detecting *a priori* unknown structure in data. This is, indeed, not a new problem. A vast literature in statistics, optimization, control, prediction, and information theory, addresses itself to problems of this nature. In fact, the use of *state space* methods has recently come to the fore in times series analysis.[8] What distinguishes our work is the incorporation of concepts from dynamical systems theory:

(1) the notion of global stability (attractors);

(2) the deterministic production of apparently random behavior (chaos);

(3) quantitative measures of temporal complexity (metric entropy and Lyapunov characteristic exponents);

(4) the notion that relevant state space coordinates can be developed systematically from a data set (reconstruction); and, finally,

(5) the consideration of manifestly nonlinear behavior.

The approach outlined here proposes a set of problems through which we may complete the line of investigation concerning the geometric characterization of apparently random behavior that has developed over the last half dozen years. Going beyond this, it suggests an approach that employs global structure in chaotic data analysis. Previously, measures of chaos have been based on averages or random samples of local structure. When employed in real data analysis these have often suffered from large errors[9] and from spurious results[10] and also in their consistent interpretation.[10, 11] This is due largely to the omission of global information that is manifestly contained in the data.

Before presenting the method for deducing deterministic equations of motion, we will first briefly review chaotic data analysis and attractor reconstruction. Following the method's description we show how it leads to an estimate of the minimum embedding dimension, and to estimates of the Lyapunov spectrum, metric entropy, and extrinsic noise level. The analysis of data from several numerical and experimental examples then illustrates its application. We conclude with a few comments on implementation and a discussion of applications to other more complex dynamical systems, to prediction and control systems, and to scientific model building.

2. Data acquisition

We must first describe the nature of the information with which data analysis begins. A data series is a set of N sequential temporal data points $\{\vec{v}_n \colon n = 0, \cdots, N-1\}$. The sampled data is obtained every sampling interval τ_s with measurement resolution ϵ_s. Information is acquired at the measurement channel rate $C_{acq} = -\,\tau_s^{-1} \log_2(\epsilon_s)$: the communication capacity[12] of the measurement channel.[7] This rate imposes an upper limit on the observable complexity of the process which can be entirely reconstructed. Specifically, the measured metric entropy h_μ is bounded: $h_\mu \leq C_{acq}$. In the typical experimental situation this upper bound is rarely approached. In other words, measurements of the state variables contain vastly more information than the dynamics which generated the data series is capable of producing. To illustrate the basic method we will be concerned with time series: a temporal sequence of a single experimental observable.

The overall approach to chaotic data analysis that we present here consists of five parts. These are not necessarily separate steps, e.g. some estimates can be improved iteratively. First, one chooses a reconstruction technique and transforms the data into the state space. Second, the dimension of the reconstructed data is estimated to provide an initial guess of an upper bound on the embedding dimension m_{bed}. Third, the equations of motion are estimated; this also yields the minimum embedding dimension. Fourth, a number of related statistical quantities are computed, such as a estimates of the extrinsic noise $\vec{\sigma}_{ext}$, the Lyapunov spectrum, information dimension, and metric entropy. Finally, these steps can be repeated with data from successive control settings to get an arc in $\mathbf{D}_M$ of a parametrized family of dynamical systems.

3. Reconstruction

A reconstruction technique[4] R is a nonlinear (diffeomorphic[6]) coordinate change from $\vec{v}$, the sampled data, to $\vec{x}$, the reconstructed state space coordinates. Several reconstruction techniques have been used to date. Consider a single continuous-time scalar signal $v(t)$. In the derivative method the coordinates are developed from the signal as successive temporal derivatives: $\vec{x}(t) = (v(t), \dot{v}(t), \ddot{v}(t), \cdots)$. The most widely-

---

with spatially-extended or fully-turbulent systems this may not be the case. The dominant problem often becomes how to obtain sufficient information to characterize spatial pattern evolution.

used method is delay reconstruction. Here the coordinates are taken as successive delays of a signal, $\vec{x} = (v(t), v(t-\tau), v(t-2\tau), \cdots)$. Spatially-separated probes have also been used as state space coordinates: $\vec{x} = (v_x(t), v_{x+\delta}(t), v_{x+2\delta}(t), \cdots)$. τ and δ are free parameters chosen to yield optimal reconstructions; where optimality is determined by the application. The Karhunen-Loéve transformation applied to these coordinates greatly reduces a reconstruction's sensitivity to τ and δ.[13] We note in passing that all of these techniques can be found in one form or another in the above-mentioned body of related literature as scatter plots, time-differenced series, contiguity lag models, multivariate time series, and so on. This literature apparently does not supply, however, a systematic theory to aid in the selection between the methods nor is there any geometric interpretation of reconstruction itself like that provided by dynamical systems theory.

As will be shown, deducing the optimal equations of motion leads to an estimate of the *minimum* embedding dimension. To get started, however, an estimate of an upper bound for m_{bed} is helpful. This can be found using standard techniques to estimate the local dimension m_{local}.[5, 14, 15] With this, the initial guess, an upper bound, for the embedding dimension is $2m_{local}$.[16]

The reconstruction method R can introduce unknown distortions that complicate the representation of the dynamic in some given function basis. Ultimately, one would like to generalize the optimization method described below to search in space of reconstruction techniques, i.e. embeddings, and also to look for nonlinear coordinates. For our present purposes, we shall assume that an adequate reconstruction is available. With this, the data can be embedded in a hierarchy of state spaces of increasing dimension.

4. The EoM procedure

By definition the function basis $\vec{\phi}$ spans the model space $\mathbf{D}_M$: $\vec{F}(\vec{x}) = A\vec{\phi}(\vec{x})$. We assume that this can be done systematically for any m_{bed}. With any such basis the task reduces to the common statistical problem of estimating the parameters $\{a_{ik}\}$, $A \in \mathbb{R}^{\mathrm{K}m_{bed}}$, from a cloud of data points that lie on or near the deterministic dynamic. Numerous expansion bases are available for this, such as Taylor, Chebeysev, rational functions, and splines. One might even consider expansions in which the coefficients entered as nonlinear parameters. Although these choices are important for a particular application, they do not affect our method's overall implementation, only the ease and accuracy with which equations of motion are obtained and the apparent simplicity, or lack thereof, of the estimated equations of motion themselves.

Roughly speaking, there are two estimation classes: one based on using data from the entire attractor to estimate parameters; the other based on using data from local regions on the attractor. The first we call the **global EoM** procedure; the latter, the **atlas EoM** procedure. We shall concentrate on the former, as it is simplest to describe and leads to equations of motion most like those with which we are familiar. The atlas method, as we shall describe below, is closely allied to differential topology[17] and is more general than the global approach in the sense that fewer statistical and geometric assumptions about the data are required. Consequently, the atlas approach can be successfully applied to a wider class of behavior.

The goal in estimating the deterministic equations of motion from a noisy data set is to deduce a minimal model that reproduces the behavior. To do this we first need a measure of deviation of the data from a given dynamic $\vec{F}$. The *observed noise* $\vec{\sigma}_{obs} = (\sigma_{obs,0}, \sigma_{obs,1}, \cdots, \sigma_{obs,m_{bed}-1})$ provides this and is defined component-wise by

$$\sigma^2_{obs,i} = \frac{1}{N} \sum_{n=0}^{N-1} (y_{n,i} - F_i(\vec{x}_n))^2 \tag{4.1}$$

where $\vec{y}_n = (y_{n,0}, y_{n,1}, \cdots, y_{n,m_{bed}-1})$ is the state that succeeds $\vec{x}_n$ in the data set. This measures the error in predicting the observed next state $\vec{y}_n$ using the estimated dynamic on the observed current state $\vec{x}_n$: $\vec{F}(\vec{x}_n)$. In the statistics literature it is called the "one-step prediction error variance".

The second requirement is a goodness-of-fit measure that reconciles the two conflicting tendencies of the improvement in fit and the model's increased complexity, with increasing approximation order.[18] This is used to objectively select an optimal model from the range of those consistent with the data. It is, in fact, a "cost function" that ranks model candidates in $\mathbf{D}_M$. The *information contained in a model* or *model entropy* $I(M)$ is given, in a simple approximation, by

$$I(M) \simeq \sum_{i=1}^{m_{bed}} \log_2 \sigma_{obs,i} + \sum_{\substack{i=1\\k=0}}^{m_{bed},\,T(K)} \log_2 \sigma_{a_{ik}} \simeq \log_2 \sigma_{obs} + m_{bed}\, T(K) \;, \tag{4.2}$$

where $T(K) = \dfrac{(m_{bed}+K)!}{m_{bed}!K!}$ is the number of basis functions up to order K in m_{bed} variables, $\sigma_{a_{ik}}$ are the error variances of the parameters, and $\sigma_{obs} = \parallel \vec{\sigma}_{obs} \parallel / m_{bed}$. The basic principle for selecting the optimal model in $\mathbf{D}_M$, an informational Occam's razor, is that the model entropy $I(M)$ is minimized. This should be compared with the maximum entropy formalism of Jaynes.[19] We could have simply selected *ad hoc* a particular approximation scheme, such as linear or quadratic Taylor functions. With such a restriction, the model entropy does not play an important role. However, the inclusion of the optimality criterion of minimum model entropy allows for the procedure itself to select the most appropriate scheme.

As defined below, $I(M)$ differs from conventional "model identification criteria"[20] in that it accounts for the effect of deterministic amplification or damping of $\vec{\sigma}_{ext}$. This is especially important for time- or space-dependent data series derived from locally unstable dynamics. Any such criterion applied to data produced by a chaotic dynamical system, for example, that ignores this systematically underestimates the "goodness of fit" and reports that the data is random when it may be wholly deterministic.

---

m_{bed} dependence of $T(K)$ for notational convenience.

A few comments on $I(M)$ are necessary to suggest its interpretation and how the above approximation is developed. Consider an ensemble of experiments each of which produces a data set and, upon EoM analysis, a model. The ensemble of models is described in part by the distribution $P(M)$ on $\mathbf{D}_M$ of estimated dynamics. The model entropy is formally

$$I(M, M') = \int_{\mathbf{D}_M} dm\ P(m) \log_2 \frac{P(m)}{P(M')} \ , \tag{4.3}$$

where $P(M')$ describes the *a priori* distribution of models M' in the class specified by Eqs. (1.1) and (1.2). Recalling the Gaussian approximation that is implied there and noting that $P(M)$ consists of two components: deviations due to errors in parameter values and "errors" due to $\vec{\sigma}_{ext}$ for a given order K expansion, the model entropy is

$$I(M) = \log_2 \frac{\sigma_{obs}}{\sigma_{ext}} + \frac{\sigma_{ext}^2 - \sigma_{obs}^2}{2\sigma_{obs}^2} + \frac{1}{2N}\left(\sum_{n=0}^{N-1} \frac{(\vec{y}_n - \vec{F}(\vec{x}_n))^2}{\sigma_{obs}^2} + T(K)\right) , \tag{4.4}$$

where $\sigma_{ext} = \|\,\vec{\sigma}_{ext}\,\|/m_{bed}$. Ignoring the deterministic amplification of $\vec{\sigma}_{ext}$, i.e. the first and second terms, using a subjective estimate of the measurement errors for $\vec{\sigma}_{obs}$ in the third term, and dropping common, constant factors, we recover in the remaining third and fourth terms an existing information criterion based on maximum likelihood.[20]

5. Lyapunov characteristic exponent spectrum

The full spectrum of m_{bed} Lyapunov exponents $\{\lambda_i : i = 0, \cdots, m_{bed}-1\}$ may be estimated using the optimum model. The Lyapunov exponents associated with directions transverse to the orbit and off the attractor, i.e. not directly constrained by information in the data set, are estimated in the sense of optimality defined by $I(M)$. The vector field not constrained by the data set is informationally the simplest with respect to the chosen reconstruction R and function basis $\vec{\phi}$.

The procedure for this estimation derives directly from the Lyapunov characteristic exponent (LCE) spectrum's definition. For discrete-time dynamics (or discretely-sampled continuous data)

$$\lambda_i = \int d\mu(\vec{x}) \log_2 \|\,\partial\vec{F}^n(\vec{x}) \bullet \vec{e}_i\,\| \ , \tag{5.1}$$

where $\vec{e}_i$ is the i$^{\text{th}}$ basis vector in the tangent space; the index i is taken so that the spectrum is monotonically decreasing; $\partial\vec{F}$ is the matrix of partial derivatives; and $\mu(\vec{x})$ is the invariant measure on the attractor, which is simply estimated by the data itself. This yields the LCE estimator given by

$$\lambda_i \underset{N\to\text{large}}{\simeq} \frac{1}{N}\sum_{n=1}^{N} \log_2 \|\partial\vec{F}(\vec{x}_n) \bullet \vec{e}_i\,\| \ . \tag{5.2}$$

Thus, the computational technique is straightforward and directly related to that developed for numerical simulations.[21, 22] The overall approach to estimating the full Lyapunov spectrum proposed here should be compared with estimations based on averages of linear approximation to the tangent space from local data.[9, 10, 11]

The metric entropy is then estimated as the sum of positive λ_i; the information dimension, via the Kaplan-Yorke formula.[23] We note that another independent estimate of the metric entropy may be developed from the local spreading due to $\vec{F}$ as measured by a conditional probability density.[24, 25] This estimator forms the basis of optimal prediction algorithms described in the applications at the end.

6. Extrinsic noise level

If, as we have assumed, the data is produced by a chaotic system the observed noise level $\vec{\sigma}_{obs}$ measures, but does not distinguish, two sources of "error": those due to the deterministic amplification of extrinsic fluctuations and the fluctuations themselves $\vec{\xi}$. This decomposition is expressed informationally as

$$\log_2 \sigma_{obs} = h_\mu + \log_2 \sigma_{ext} \ . \tag{5.3}$$

The units for each term are bits per time unit; where the time unit is τ_s.

Once the deterministic amplification of extrinsic noise is estimated by the metric entropy or Lyapunov spectrum, an estimate of the "true" extrinsic noise level follows

$$\sigma_{ext} \simeq \sigma_{obs}\ 2^{-h_\mu} \ . \tag{5.4}$$

With respect to the assumed model class, this yields an estimation of the fluctuations actually present in the system and not generated by the deterministic dynamics. The methods of estimating $\vec{\sigma}_{ext}$ used in references [25] and [7] should be compared to this. In this context, $\log_2 \sigma_{ext}$ plays the role of a thermodynamic entropy: the information missing from the deduced "macroscopic" variables and parameters of the deterministic dynamic. We shall return to this thermodynamic interpretation in a later section. To summarize, there is no need for an *a priori* estimate of the extrinsic noise level, as one often finds explicitly or implicitly in current chaotic quantifiers. We have replaced this with the model hypothesis of Eqs. (1.1) and (1.2) and model entropy minimization.

7. The attractive hypothesis

A further important assumption is that the reconstructed data lies near or evolves to an attractor. Although this need not be the case if the EoM analysis is being applied in a context in which explicitly transient, globally unstable behavior has generated the data set, it is worth discussing in some detail for several reasons. First, it is the main case of interest to us here; there are several interesting applications. Second, it aids considerably in improving estimated equations of motion and in reducing computational resources. And third, it is an often unspoken assumption that deserves explicit acknowledgement in any statistical analysis of chaotic data. It articulates one's anticipation of an important property of the data, that may or may not be born out during analysis. As such, we call it the *attractive hypothesis*.

The mathematical statement formalizes the requirement that the data "map" into itself. We define the state space volume $\Lambda_{attract}$ as the largest Euclidean m_{bed}-cube circumscribing the reconstructed data points $\{\vec{x}_n\}$:

$$\Lambda_{attract} \equiv \{ \, \vec{x} \in \mathrm{I\!R}^{m_{bed}} : \min \{x_{ni}\} \leq x_i \leq \max \{ \, x_{ni} \} \, , \, i = 0, \cdots, m_{bed} - 1 \, \} \; . \quad (5.5)$$

The extrema are taken over the entire data set; the subscript i refers to the i^{th} coordinate. With this, the attractive hypothesis requires

$$\vec{F}(\Lambda_{attract}) \subseteq \Lambda_{attract} \; . \quad (5.6)$$

For brevity's sake and as it does not affect the main points of our discussion, we consider only the deterministic dynamic and not the model's stochastic component.* This particular statement of the attractive hypothesis is rather strong. A weaker form, for example, would be to limit the constraint to a ϵ-cover of the data set. But this has its own difficulties and so we shall use the above form.

The geometric interpretation of the attractive hypothesis is rather simple for continuous time flows, Eq. (1.1): the vector field sufficiently far from the data cloud points inwards. Thus, any simulated trajectory starting outside the region constrained by the data will move toward $\Lambda_{attract}$. For discrete time maps the geometric picture is simply that an initial condition maps into $\Lambda_{attract}$ on the first iteration.

In implementing the attractive hypothesis, the vector field or discrete-time dynamic outside the data-containing region is augmented with a smooth contractive extrapolation. To facilitate this the data is normalized initially to $[-1, 1]^{m_{bed}}$. Then one of several standard extrapolations outside this cube are imposed. One alternative is to simply add data outside the cube that provides the desired extrapolation. Another alternative is to employ a basis that explicitly implements the attractive property outside the cube, e.g. Hermite polynomials rather than Taylor functions.

As discussed in this section the attractive hypothesis applies to either global or atlas EoM analysis. It imposes boundary conditions on the chosen fitting procedure, toward whose implementation details we now turn.

8. Dynamic estimation and a diagnostic

This completes the theoretical background of our method. In this section we describe procedures for fitting the *graph* of the dynamic $\vec{F}$ or, equivalently, estimating the parameters $\{a_{ik}\}$. There is a large number of algorithms for the central fitting and optimization task that we have just introduced: least squares, simplex methods, simulated annealing, and the Boltzmann machine, come to mind. After experience with the first three, we have settled on the singular value decomposition implementation of least squares fitting.[27, 28] This is very robust and stable for the type of over-

---

*Ruelle[26] provides a detailed discussion of attractors in the presence of fluctuations.

determined and nearly singular problems with which we are concerned here.[†] Furthermore, it does not require an orthogonal basis as errors in parameter estimates are independent due to the implicit use of Karhunen-Loéve transformation.

There are a number of practical concerns regarding the choice of basis functions. For example, if we use smooth basis functions to fit a dynamic with apparent (e.g. circle map) or real discontinuities (e.g. Lorenz map) or that is not smooth (e.g. tent map), the fits are naturally very poor. Similarly, a term in the real dynamic may not be optimally approximated in the chosen basis: e.g. a periodic driving term approximated by a Taylor expansion. A basis may be appropriate if the data is properly transformed, however. For example, one can use $z = sin(wt)$ rather than t as a coordinate for a Taylor function basis in the case of data from a driven oscillator.

To aid in the detection of poor approximation, we have developed a simple diagnostic. Although it does not take into account particular features of a chosen basis, it does provide a measure of the difficulty encountered using smooth bases to fit a dynamic whose graph has high state-space frequency components. We use the ϵ-convergence properties of a dimension-like quantity that measures the RMS range of the dynamic over ϵ-size domain cells. Its convergence is monitored in the same manner used for entropy convergence to investigate the effects of noise on symbolic dynamics.[25] The diagnostic is interpreted as follows. If the convergence is slow, there is much variation in the dynamics and the estimation will require high order or fail altogether. If the convergence is rapid, then a smooth approximation is likely to work.

Ultimately, we believe a procedure will be found that avoids this class of fitting problem altogether by determining the optimal nonlinear basis directly from the data set. It would then go on to estimate the equations of motion in a form compact with respect to that basis. A first step in this direction is the atlas EoM procedure presented in the next section.

9. Atlas EoM estimation

To balance these cautionary remarks on implementation details, we digress at this point to indicate the general applicability of deducing equations of motion. The proof that EoM analysis always works *in principle* for smooth dynamical systems relies on the Morse lemma. This states that any smooth manifold is approximated by a set, or *atlas*, of local quadratic polynomial *charts*.[29] In the case of EoM analysis, there are two manifolds of interest: (i) the reconstructed state space, the domain of the dynamic, and (ii) the graph of the dynamic. If, as we have assumed, the underlying dynamical system is smooth, there is a finite domain cell size ϵ below which the coordinate charts for the dynamic allow adequate approximation within some error level δ in the range of the dynamic. With real data, the minimum ϵ is bounded from below by $\vec{\sigma}_{ext}$. It may be

[†]This is an optimization problem with the cost function being $I(M)$. We have implemented simulated annealing and a simplex method. For low dimensions ($m_{bed} < 5$) the singular value decomposition is much faster. Perhaps for larger m_{bed}, where the computational resources for singular value decomposition are prohibitive, these alternative techniques will be preferred.

much larger, however, indicating a simpler state space manifold and dynamic and that fewer than $O(\vec{\sigma}_{ext}^{-m_{bed}})$ charts are necessary.

Employing this topological framework, our atlas EoM procedure computes a global list of coordinate patches to the graph of the dynamic, approximating the dynamic with a local spline and noting the domain of applicability for each. The splines may go up to cubic order in our implementation. Thus for a fixed error δ, the domain cell size is chosen to be $\epsilon \simeq \delta^{-k}$ for an order k spline. Within each chart the spline parameters are estimated from the local data using singular value decomposition.* A number of splines are available for this task: B-spline, Bezier, and Hermite. Each provides its own type of local approximation. We have implemented B-splines as they allow both the first and second derivatives to be continuous across the charts. This results in a globally C^2 dynamic. We note that B-splines also allow for rapid computation of the LCE spectrum. Finally, the model entropy is simply summed over the fit of each local chart.

The atlas procedure is tantamount to implementing EoM analysis with a function basis of splines. This approach allows a much wider range of dynamics to be fit than by global fitting with conventional bases. Unfortunately, its implementation is more difficult and it produces "piecewise" equations of motion. The latter are not what one typically considers a "closed" form model. From the computational point of view, however, the complexity in atlas equations of motion is only apparent. Theoretically, the vast reduction in algorithmic complexity of the chaotic data to a dynamical system is the same in the atlas procedure as for the global method. Properly speaking, the algorithmic complexity is the same for both methods; although their required computational resources differ by some constant factor. Practically, with our atlas-based dynamical system simulator, the difference is noticed only as a moderate slow down in simulation speed.

Local approximation of the dynamic over small regions in the state space has also been discussed previously in the time series literature by Priestley,[31] under the name of *locally-linear autoregressive moving average* (ARMA) *models*. In these the dynamic is approximated by simple linear regression or interpolation. The central motivation is forecasting time series. Independently, references [32] and [33] have suggested very recently the same approach. The former provides a wide-ranging review of efficient data structures and algorithms for artificial intelligence computation. A hierarchical tree-structured piece-wise linear atlas is proposed as an efficient method for learning behavior and also for sequence prediction. Local data is linearly interpolated to produce the predictions. The second reference evaluates this approach for predicting chaotic time series. This work does not discuss, however, deducing equations of motion, simulating them, or other important facets, such as the attractive hypothesis, extrinsic and

*When applied to the data in the domain of each chart, singular value decomposition also provides a method for simultaneously estimating m_{local}. This is estimated as the rank of the local covariance matrix, computed by the number of significant singular values, averaged over each chart in the atlas. This method is closely related to the "local linear regression" method of dimension estimation.[14, 30]

observed noise, *etc*....

There has also been much related work in chaotic data analysis that examines statistics from local linear fits. The first attempt to measure chaotic attractor dimensions employed a *local linear regression* to obtain a piecewise linear approximation to an attractor.[14] More recently, work on measuring the Lyapunov characteristic exponents from chaotic data has employed local estimates of the tangent mapping, which is itself a locally linear approximation of the dynamic.[11, 34]

In our own investigations of piecewise linear atlases we have found them to be unreliable indicators of the underlying *deterministic* behavior. Simple examples, even in one dimension, show that piecewise linear equations of motion can exhibit periodic behavior when the original dynamics is chaotic and *visa versa*. This depends on the chart size and on translation of the atlas spline knots as a whole. The other consideration which has lead us away from piecewise linear dynamics to smoother atlases is that they violate the physically-motivated hypothesis of smooth dynamical systems. Most physical processes do not exhibit arbitrarily fast changes in their first derivatives. Finally, without smooth continuation between charts, it is not clear how far differential topology can be applied to piecewise linear equations of motion.

In their defense, there are some data sets for which abrupt changes in derivative and even discontinuities are appropriate. Additionally, if the chart size is near $\vec{\sigma}_{ext}$ and this noise is added during a simulation, then piecewise linear atlases will exhibit behavior that is noise-equivalent to a smooth atlas.

We shall not discuss the atlas procedure further due to the substantial complication of the method. In an effort to convey the central ideas of EoM analysis, in the remaining sections we will consider only the global EoM procedure: global function fitting over the reconstructed attractor.

10. Global EoM estimation

The global EoM procedure operates component-wise to estimate each f_i in turn. The result is the estimated dynamic $\vec{F}$, a set of m_{bed} functions, and m_{bed} noise levels. The basic method is to compute $I(M)$ as a function of approximation order K and m_{bed} and also as a function of the removal of particular basis functions. Additionally, the singular value decomposition computes singular values that measure the parameter error ellipsoid size, and these too may be reduced in number in order to minimize $I(M)$ and so improve the fit at each step. The latter is taken as the inner most loop in our procedure, then the order K is varied, and finally, the embedding dimension is changed. We start m_{bed} at $2m_{local}$ and K at a large value permitted by the computational resources. These are reduced by removing coordinates and basis functions which decrease $I(M)$ until there is no further improvement. The resulting m_{bed} is the minimum embedding dimension.

Once the optimum model is obtained in this way, then it may be compared to the original data via numerical simulation of Eqs. (1.1) and (1.2).* The comparison requires that the appropriate amount, $\vec{\sigma}_{ext}$, of noise be added to the simulation and that any coordinate transformation, due to reconstruction or later modification be inverted.

For data series from continuous-time signals, the integral form of Eq. (1.1) is estimated. That is, we approximate the *flow* $\phi_t\colon \mathbf{M} \to \mathbf{M}$, where $\vec{x}(t) = \phi_T\,(\vec{x}(t-T))$, and ϕ_T is a discrete-time mapping given by

$$\vec{x}(t) = \vec{x}(t-T) + \int_{t-T}^{t} d\tau\ \vec{F}(\vec{x}(\tau)) \ , \tag{10.1}$$

There are two benefits of fitting this integral form over fitting Eq. (1.1) directly. First, more robust estimates of the parameters are obtained due to the averaging of short-time scale noise. Second, the same fitting algorithm may be used for discrete and continuous time data streams. When the signal is fit in this integral form our method of estimating the parameters of $\vec{F}$ is highly reminiscent of linear prediction filters, used extensively in speech synthesis, for example. In linear prediction coding, the f_i are linear. In contrast, our method, via Eq. (10.1), constitutes a nonlinear prediction filter. We shall return to this topic in the last sections.

There are four general comments on the method's implementation. First, although we have not done an exhaustive study of data set size scaling, experience indicates that a relatively small number of points, less than a few thousand, is adequate for chaotic attractors in up to 4 dimensions.

Second, in comparing original to model orbits, one must keep in mind that the comparison is both geometrical and statistical. A corresponding level of noise $\vec{\sigma}_{ext}$ must be added during the simulation of the estimated model. If the behavior is chaotic, the model does *not* reproduce the original data series nor is $\vec{F}$ itself necessarily the "true" deterministic dynamics and $\vec{\sigma}_{ext}$ the "true" extrinsic fluctuation strength. The method makes three claims: (i) The original dynamics and the estimated model are in the same noise-equivalence class. (ii) Almost all simulated orbits are statistically similar, in the sense of (say) power spectra and statistical moments, to the orbit reconstructed from the original data. And, (iii) almost all simulated orbits lie on a branched manifold that is topologically equivalent to the reconstructed attractor.

Third, for the global EoM procedure, the required computational resources, that is, memory and time, scale as $N^2 m_{bed} T(K)$. The number of fit parameters must also be kept smaller than the number of data points. Some authors suggest the upper bound for this is $\sqrt{N}$. We note that the method is easily parallelized for distributed processors or adapted to array processors, and so may be greatly speeded up. Even so, on current

*As a necessary tool in EoM analysis, we have written an interactive simulator that directly accepts a basis type, such as Chebeysev, Taylor, or B-spline, and a list of expansion coefficients produced by the EoM analysis.

Figure 1.

Effect of extrinsic noise on model entropy. $I(M)$ versus expansion order $K \in [0, 11]$ and extrinsic noise level $\vec{\sigma}_{ext}$ for the one-dimensional map described in the text. The extrinsic noise level was 10^{-18}, 10^{-14}, 10^{-10}, and 10^{-6}, for the stars, circles, squares, and diamonds, respectively. At each noise level, the Taylor coefficients are reconstructed with errors less than 10^{-3}.

scientific workstations we expect through further optimizations to estimate global equations of motion in up to 8 variables using function bases up to fourth order.

Finally, we note that transient data and data from different experimental runs can be used. Indeed, in some circumstances this can greatly improve the accuracy of EoM analysis. This is because data off of the attractor serves to constrain parameter estimate in regions of the dynamic which otherwise would not be determined directly.

11. Numerical examples

We will illustrate EoM analysis with the global procedure using a Taylor function basis for three numerical and three experimental chaotic data sets. Although extensive testing has been performed, only these examples will be described as they illustrate not only that the method is readily applied to real data, but also some limitations and several additional features whose more general discussion cannot be included. Unless otherwise noted, all examples are discrete time maps in low dimensional systems. Another motivation for the choice of examples is that with data from a low-dimensional system several of the important features of the analysis can be graphically demonstrated. Results for high dimensional systems do not admit simple illustration.

The first example shows the effect of extrinsic noise on the model entropy and the optimum model. We consider the stochastic logistic map $x_{n+1} = rx_n(1 - x_n) + \xi_n$ where the nonlinearity parameter r is 3.7, so that the deterministic behavior is chaotic. ξ_n denotes uniformly distributed noise with amplitude σ_{ext}. Figure 1 shows the dependence of $I(M)$ on expansion order K and noise level σ_{ext}. Note that T(K) = K for one dimension. 1000 iterates were used in the analysis.

At each noise level, the parameter r, as well as the other implicit parameters, are correctly estimated to within .1% error or better. The extrinsic noise level is estimated with similar accuracy. The overall feature to note in the figure is a distinct minimum in the model entropy that identifies the optimum model. Initially with increasing approximation order, the model entropy drops steeply to a *convergence floor* once a sufficient number of terms are reached. In this example, this occurs at the inclusion of quadratic terms. Beyond this there is a linear increase as the model grows in complexity with added redundant terms.

The next figure demonstrates the estimation of the minimum embedding dimension in terms of the trade off between m_{bed} and approximation order $T(K)$. Figure 2 shows a perspective plot of the model entropy surface for the logistic map, just described. The reconstruction used delayed coordinates in up to nine embedding dimensions. The approximation order went up to 25. Although difficult to discern in the plot, there is a gradual increase in $I(M)$ with m_{bed} along the trough. This increase is sufficient to select the $m_{bed} = 1$ model as optimum.

The second numerical example used the two-dimensional Hénon map to generate a time series:

$$\begin{aligned} x_{n+1} &= y_n + 1. - 1.4 x_n \\ y_{n+1} &= .3 x_n \end{aligned} . \tag{11.1}$$

The reconstructed attractor in various dimensions was obtained using successive delays of x_n. Figure 3 shows $I(M)$ as a function of m_{bed} and $T(K)$ for this example. 1000 points were used. The optimum model, indicated by the dot in the figure, was found to be

$$\begin{aligned} x_{n+1} &= .839 - 1.76 y_n \\ y_{n+1} &= .785 + .016 x_n + .300 y_n - 1.79 x_n^2 \end{aligned} . \tag{11.2}$$

(Terms with parameters smaller than 10^{-3} have not been included.) Here, the variables x_n and y_n refer to the original variable x_n and its first delay x_{n+1}. There is no discernible difference between the original reconstructed attractor and the one simulated

Figure 2.

Estimating the minimum embedding dimension. $I(M)$ versus expansion order $K \in [0, 25]$ and embedding dimension $m_{bed} \in [1, 9]$ for the delay-reconstructed logistic map, as described in the text. The minimum $I(M)$ identifies the optimum model as indicated. The extrinsic noise is zero. The Taylor coefficients for the optimum model are estimated to better than 10^{-6}.

using the optimum equations.

The third example demonstrates the type of dynamic with which global function fitting has difficulty. Two dimensional data was taken from the phase$-\dfrac{\pi}{2}$ Poincaré section of the driven, damped Duffing oscillator

$$\begin{aligned} \dot{x} &= y \\ \dot{y} &= -\gamma y + ax - bx^3 + F cos(\omega t) \end{aligned} \tag{11.3}$$

where $(\gamma, a, b, F, \omega) = (.03, 1., 10., .65, .93)$. The position and velocity were sampled at the driving period.

Figure 3.

$I(M)$ versus expansion order K and embedding dimension m_{bed} for the delay-reconstructed Hénon map, as described in the text. The minimum $I(M)$ identifies the optimum model as indicated. This model reproduces the qualitative features of the original data. The floor to the right indicates various (m_{bed}, K) for which $I(M)$ was not computed.

Upon EoM global analysis, fitting to high order exhibited only a very slow convergence and no convergence floor in $I(M)$. The diagnostic was consistent with the interpretation of a complex dynamic. This was confirmed, as shown in figure 4, with a stereo plot of the one component of the graph of the dynamic: x_{n+1} is plotted as a function of (x_n, y_n). There are sharp ridges and steep cliffs that are obviously difficult to approximate with Taylor, or any other known, function basis.*

*This is a counterexample to the statement in reference [32] that "most mappings that arise in practice have slowly varying slope".

Driven systems do lead to one interesting feature, and this will be our only direct comment on numerical investigation of continuous-time flows. Consider a two-dimensional driven oscillator, like the Duffing oscillator just mentioned. The state space is three-dimensional, consisting of position, velocity, and driving phase coordinates. When the EoM global analysis is carried out in $m_{bed} = 3$ dimensions, the $sin(\omega t)$ term must be approximated directly in the chosen basis. This leads to slow convergence and relatively poor approximation. However, when the EoM analysis is carried out in higher dimensions with the phase variable included as one coordinate among others, a Hamiltonian subsystem is detected instead of a periodic driving term. The latter could have been approximated by a series expansion, as was done in three dimensions. This would have been a more complex approximation. Instead, the procedure "chose" to substitute a dynamical subsystem to generate sin(ωt). This achieves a simplification of the optimum model by substituting a computational procedure $\{\dot{x} = y, \, \dot{y} = -\,\omega x\}$ in the place of explicit inclusion of the $sin(\omega t)$ expansion.

To conclude the numerical example section, we note that we have tested the global method using a number of other numerically-generated data sets from well-known dynamical systems. The three dimensional flows included Rössler, Lorenz, and other parameter regimes of the driven, damped Duffing, and van der Pol ordinary differential equations. Four dimensional flows included Rössler's hyperchaos[35] and the Hénon-Heiles system.[36] The latter is a Hamiltonian flow. One six dimensional flow was tested: a coupled pair of Rössler oscillators. The discrete time maps included several in each of one, two, three, four, and fifteen dimensions. The latter was data generated from a 15 site logistic lattice.[37] Reconstruction techniques used include the original variables and derivatives of single variables. To test the sensitivity to possible distortions of the data, additional transformations were applied to the reconstructed data for the lower dimension examples. These included shearing, scaling, rotation, and translation. The results were all consistent with the procedures we have just outlined. The fits were typically very good. In a few cases, such as Rössler's towel map, equations simpler than the original system were recovered.

Interestingly, time delay reconstruction, a very popular method, often leads to poor approximations. The general indication is that it produces fairly complex, difficult-to-fit dynamics. Fortunately, the atlas method gives good results with time delay reconstruction.

12. Experimental examples

One of the authors carried out[38] a limited version of the procedures outlined above a number of years ago on an electronic circuit implementation of Shaw's variant of the chaotic driven van der Pol oscillator:[39, 40]

$$\begin{aligned} \dot{x} &= y + Fcos(\omega t) \\ \dot{y} &= -\,\mu y(a - x^2) - kx \end{aligned} \quad , \tag{12.1}$$

where $(\,a, k, \mu, F, \omega) = (.11,\ .72,\ 9.,\ .23,\ 1.35)$. The chaotic attractor here consists of two "bands" or ribbons with a single folding process. There is no visible fractal structure, due to high effective dissipation. The driving phase $\phi = 0$ data from x(t) and

Figure 4.

Graph of the dynamic for the Duffing oscillator Poincaré map. A stereo pair of perspective plots of the x-component of the dynamic. The long vertical axis is $x_n \in [-.68, 70]$; the horizontal, $y_n \in [-1.17, .85]$; and the third, short axis is $x_{n+1} \in [-.68, 70]$. The latter should be viewed as coming in or out of the page, depending on one's stereo visualization method. Sharp ridges and steep cliffs are evident, indicating difficult approximation. Approximately 4000 points are shown.

y(t) was sampled approximately every 40 milliseconds by a microcomputer with 12 bits of resolution. 2048 two-dimensional points were accumulated, so that there were 1024 points on one of the bands. This band was parametrized by angle of return θ_n to yield a one-dimensional map $\theta_{n+1} = f(\theta_n)$. The function $f(\theta)$ was fit by a fourth order polynomial: $f(x) = a + bx + cx^2 + dx^3 + ex^4$, where $(a, b, c, d, e) =$ (.66, 1.9, -2.4, -1.1, .91). This equation of motion reproduced the observed chaotic behavior quite well, as expected. The single Lyapunov characteristic exponent for the map $\lambda_{map} \simeq .6$ bits per iteration was computed by Eq. (5.2) above. From that, the maximum Lyapunov exponent for the ODE was estimated: $\lambda_{vdp} = \dfrac{\lambda_{map}\omega}{4\pi} \simeq .065$ bits per time unit. The observed noise level $\vec{\sigma}_{obs}$ was approximately 2^{-9}. The extrinsic noise level was estimated then to be slightly above the measurement resolution: $\sigma_{ext} \simeq \vec{\sigma}_{obs} 2^{-\lambda_{map}} \simeq .0012$.

To illustrate the current procedure, however, we took Poincaré section data from a recent electronic implementation of the above van der Pol oscillator in a more complex parameter regime: $(a, k, \mu, F, \omega) = (.07,\ .37,\ 7.3,\ .56,\ 1.62)$. We again attempted to obtain the Poincaré map using global EoM analysis. Although the ODEs are given explicitly, the Poincaré map equations are not known *a priori*. In such a case, successful EoM analysis can provide analytic information, the estimated equations of motion, which cannot be analytically derived.

2000 two-dimensional points were used in the analysis. A Taylor basis of varying order was used to fit the data in up to 4 embedding dimensions. The attractor was reconstructed using various combinations of the two original coordinates and their delays. The minimum model entropy of -30 bits was found using the two original coordinates and seventh order approximation. There were 20 significant parameters and so we shall not quote them here.

Figure 5 compares the original orbit with one from a simulation of the estimated equations of motion. The overall agreement is quite good. One noteworthy feature in the reconstructed dynamics is the appearance of the attractor crossing itself slightly. This indicates the estimated EoM have a certain degree of noninvertibility. This is definitely not a property of the original data as it derives from an ODE which admits only diffeomorphic Poincaré maps. This is yet another topological property of the data that could in principle be imposed during the EoM analysis via the inverse function theorem. The inclusion of this would additionally complicate the implementation as it requires the simultaneous fitting of all m_{bed} coordinates of the dynamic.* The benefits would be a reduction in the number of parameters to be fit and so an increase in overall speed and accuracy. We shall not digress further on this interesting problem.

We have also analyzed nine data sets from a chaotic dripping faucet[41] with good results. We shall mention a few of these, although the details of these analyses will appear elsewhere.

We examined the drip-interval data sets shown in reference [3] (figures (a), (b), (e), and (f) on page 55) as well as several others as yet unpublished.† Out of the nine available data sets each representing different behavior, we were able to obtain good equations of motion for five. Those for which we could not obtain equations of motion appear to be undersampled, in the sense that in any embedding dimension the dynamics is not completely determined by the reconstructed states. This is not surprising given

---

*We note that the invertibility criterion is straightforwardly implemented in the atlas procedure using the local inverse function theorem for each patch of data and each local spline approximation to the dynamic.

†Please refer to the cited figures for the following discussion. 1000 points from these data sets were used in the EoM analyses.

Figure 5.

(Left) Two-dimensional Poincaré section of the original van der Pol electronic oscillator data. 2000 points from the data set are shown. (Right) An orbit from the estimated Poincaré map. Approximately 4000 iterates are shown.

the rather simple and indirect time-interval observable used.

For the attractor shown in figure (a) we found an optimum model in one dimension with the deterministic dynamic

$$x_{n+1} = .243 + 2.81\, x_n - 3.17\, x_n^2 \ . \tag{12.2}$$

The extrinsic noise $\vec{\sigma}_{ext}$ was approximately .04 and the single LCE was approximately .7. When simulated without noise the deterministic dynamic exhibits a period four orbit. With $\vec{\sigma}_{ext}$ level noise added one finds a single noisy chaotic band similar to that seen in the original reconstructed data. In a statistical sense we can argue that the EoM analysis has revealed the dynamic to be periodic, in contrast to one's initial perception of a single chaotic band with noise added. The extrinsic noise has driven the periodic deterministic dynamics into an apparently chaotic band. This inference is also supported by the scaling theory for period-doubling cascades in the presence of fluctuations.[42]

For the attractor of figure (b), the optimum model was found in two dimensions with a third-order dynamic

$$\begin{aligned} x_{n+1} &= y_n \\ y_{n+1} &= 9.348 - 16.1432\, x_n - 26.3935\, y_n + 10.3878\, x_n^2 + 34.6795\, x_n y_n \\ &\quad + 24.3786\, y_n^2 - 3.13519\, x_n^3 - 7.14675\, x_n^2 y_n - 21.6583 x_n y_n^2 - 5.91889\, y_n^3 \end{aligned} \cdot(12.3)$$

σ_{ext} was estimated to be .01. The simulated attractor exhibited two noisy bands.

Figures (e) and (f) are examples of undersampled data sets. The "attractor" in figure (e) passes through itself in one region in up to five delay-reconstructed dimensions although it appears locally two-dimensional, $m_{local} \simeq 2$. The data in figure (f) is also similarly degenerate near the "knotted" regions with $m_{local} \simeq 2$. Embedding in higher than $2m_{local}$ dimensions does not yield a nondegenerate mapping for the estimated dynamic. From Whitney's embedding theorem, we conclude that these data sets are undersampled. In this sense they are projections from some attractor onto the inadequate coordinates provided by drip-interval sampling. To reconstruct behavior in this regime more probes, such as spatial probes across meniscus under the orifice, appear necessary.

13. Chaotic data analysis

We have outlined a general approach to estimate an "experimental" model that optimally reduces a complex data set to a compact algorithmic form. The deduction of equations of motion from data series as presented here should enjoy the same range of applicability as current statistical measures of chaos. It adds a significant and new type of qualitative information to chaotic data analysis that moves in the direction of including global structure of the underlying dynamics. Since global structure is taken into account the method offers more robust characterizations of complex data than present techniques. It also offers the future possibility of comparing "experimental" equations of motion to those derived from first principles. We anticipate that further optimizations will extend the range of usefulness to higher complexity systems. EoM analysis, as we have outlined it here, puts into a single framework a number of procedures used in chaotic data analysis: the determination of the minimum embedding dimension, attractor dimension and entropy, and the spectrum of Lyapunov characteristic exponents.

Further comment is in order to clarify the relationship between global and atlas equations of motion. A set of global equations is *always* to be preferred over an atlas model. The former (i) affords larger compression of the data into smaller algorithmic form; (ii) allows for faster computation, simulation, and prediction; and (iii) typically provides a better estimate of the vector field outside the data set. The atlas method, although more widely applicable, suffers from generating more complex models: typically, $I(M_{atlas}) > I(M_{global})$. For data that admits a global model, it is generally preferred since it represents something close to a minimum algorithmic reduction of the data with respect to the chosen function basis. Furthermore, it indicates that some property of the function basis is appropriate to representing the data. With the atlas method there is no direct indication of this additional structural information.

EoM analysis can be also used to deduce parametrized families of dynamical systems. Parameters are estimated across a family of equations of motion, each of which is obtained from data at different experimental control settings. Assuming one followed a single attractor through a bifurcation sequence, then the particular changing parameters could be identified. The smooth variation of the dynamic through a bifurcation sequence provides yet another constraint in parameter estimation and so can be used to improve the estimates at each parameter setting. At present, we have carried this procedure out only for the logistic and Hénon maps to identify the nonlinearity parameter in each.

We turn now to discuss three areas of application that motivate our longer term interest in EoM analysis.

14. Spatially-extended dynamical systems

With straightforward modifications EoM analysis can be applied to spatially-extended systems or higher-dimensional systems, such as neural networks and network dynamical systems. To focus the discussion we shall consider only spatially-extended dynamical systems in this section.[37]

In spatially-extended systems, there are other independent variables in addition to time. Associated with each we must chose an optimum reconstruction method. For example, consider space-time patterns u(x,t) produced by a partial differential equation of the form

$$\dot{u}(x,t) = F(\, u, \nabla u, \nabla^2 u, \cdots) \ . \tag{14.1}$$

Rather than estimating the differential form of spatial derivatives, it is statistically more robust to estimate the integral form

$$w_K(x,t) = \int_{-\infty}^{\infty} dy\, K(x,y)\, u(x,t) \ , \tag{14.2}$$

where $K(x,y)$ is the kernel associated with the highest spatial derivative. Thus, one fits to an partial integral equation rather than the partial differential equation. If the dynamics contains a diffusion operator (∇^2), then the kernel is the Gaussian

$$K(x,y) \ \propto \ \exp\left(\frac{-(x-y)^2}{2X^2} \right) \ . \tag{14.3}$$

Here, X is conventionally interpreted as a spatial diffusion length. For the purposes of spatio-temporal reconstruction, however, it is a free parameter analogous to T in Eq. (10.1). These considerations are similar to the preceding discussion on estimating the flow rather than the ODE directly.

Once the form of the space-time dynamic has been selected, obtaining the equations of motion for spatio-temporal data proceeds just as we have outlined in the preceding sections. The major differences are (i) that we submit space-time patches of data to analysis, (ii) the interpretation of the results as a space-time dynamic, and (iii) the type of simulator used to study the estimated equations of motion. The remaining theoretical issue for applying EoM analysis to evolving pattern data is an embedding criterion so that the local pattern dynamics is non-degenerate.[43] We now describe this is some detail for discrete-space and discrete-time *lattice dynamical systems.*[37]

Spatial and temporal entropies measure the rate of spatial and temporal decorrelation.[44] We denote these entropies h_μ^s and h_μ^t, respectively. They relate the amount of information an observer has about the edge of a pattern (spatial entropy) or about the next state (temporal entropy), given the asymptotic statistics of space-time patches.[43, 45] We assume that each probe has the same measurement resolution ϵ. With each measurement, a probe provides $-\log_2 \epsilon$ bits of information. To provide sufficient information for reconstructing the space-time dynamic, the probes must be within a space-time region **R** delimited by spatial and temporal decorrelation. We can approximate this by the space and time separation over which signals from two probes become mutually unpredictable. The region **R** is bounded then by $(\pm\tau_{\max}, \pm\,\delta_{\max})$, where $\tau_{\max} = -\dfrac{\log_2 \epsilon}{h_\mu^t}$ and $\delta_{\max} = -\dfrac{\log_2 \epsilon}{h_\mu^s}$ measure informationally the temporal and spatial decorrelation lengths.

The number of probes is determined by the complexity of the observed behavior and the requirement that the local patterns in the template lead at each time to a unique state at the spatial point of interest.* This is the criterion for *local embedding* of the spatio-temporal dynamic. Quantitatively it is measured using the *indeterminacy*[7] applied to the template data.

Consider the set of neighborhood patterns $\{nh_n^i(r)\colon n = 0, 1, 2, \cdots, i \in$ *spatial lattice*$\}$ of radius r spatial sites. There are k^{2r+1} possible neighborhood patterns, where $k = \epsilon^{-1}$ is the number of distinct measurement outcomes from a probe. A local embedding occurs when the neighborhood patterns determine the next state with probability one. That is, the conditional distribution $P(s_{n+1}^i \,|\, nh_n^i)$ is unity, where s_n^i is the measurement of the local state at time n and site i. The indeterminacy provides an informational measure of this

$$\Phi(r) = \sum_{\{nh_n^i\}} P(nh_n^i) \sum_{\{s_{n+1}^i\}} P(s_{n+1}^i \,|\, nh_n^i) \log_k P(s_{n+1}^i \,|\, nh_n^i) \;, \tag{14.4}$$

where $P(nh_n^i)$ is the observed density of neighborhood patterns $\{nh_n^i\}$. When $\Phi(r)$ vanishes then a sufficiently large neighborhood has been found with which the space-time dynamic can be reconstructed.

---

*We ignore for now the additional complication of extrinsic noise.

The number of probes N can be estimated as follows. Measurements from the set of probes must yield information at a rate higher than the total information production in the neighborhood templates. The density of information production is given by the *specific metric entropy* h_μ. The specific metric entropy is an intensive quantity with units of bits per unit volume per unit time. Within $\mathbf{R}$ the total information production is approximately $h_\mu \, \tau_{\max} \, \delta_{\max}$. The required number of probes is then

$$N \geq -\frac{h_\mu \, \tau_{\max} \, \delta_{\max}}{\log_2 \epsilon} \tag{14.5}$$

A specific space-time configuration of probes is not indicated by these considerations, however. A particular configuration is determined by the application: the neighborhood dynamics itself and the local interconnectivity variables within $\mathbf{R}$.

The occurrence of spatial amplification of noise, such as in convective instabilities, indicate the necessity of including spatial measures of perturbation propagation, such as co-moving Lyapunov characteristic exponents and multipoint entropies, dimension, and coherence.[46] Above we have seen the importance of metric entropy, dimension, and LCE spectrum and how EoM analysis can be used to estimate them. When applied to spatially-extended dynamical systems, EoM analysis will be similarly useful.

Another class of spatially-extended systems, closely allied to lattice dynamical systems, is discrete-state cellular automata. To estimate equations of motion in this case, we use Walsh functions as the function basis for $\vec{F}$. Unfortunately, with manifestly discretized state variables we cannot appeal to continuity or differentiability as constraints with which to lower the external noise level by filtering, as done in Eq. (10.1). Foregoing any filtering procedures, the first author attempted to apply this approach to estimate the neighbor transition matrix $P(s^i_{n+1} \,|\, nh^i_n)$ for the binary cellular automata-like patterns exhibited by mollusks. Digitized images of the shells* were analyzed with inconsistent results. Some of the space-time dynamics lead to propagating patterns like those found on the shells. On the whole, however, the available statistics were inadequate. This was due largely to limited data from any one shell. Additionally, the data exhibited a systematic bias due to the inherent geometric distortion of the shells' natural curvature.

On a positive note, EoM analysis has been successfully applied to numerical simulations of several lattice dynamical systems.[37] We have already mentioned above the use of a 15 site logistic lattice as a high dimensional numerical example. With the additional assumption of translation symmetry and the reduction of lattice patterns to neighborhood data, EoM analysis yields spatial equations of motion. An analogous experimental effort is underway to analyze image sequences from video feedback and several magnetic systems.

---

*The specimens included the Tent Olive, (*Oliva porphyria* Linné), the Flamed Venus (*Lioconcha castrensis* Linné), the Wavy Volute (*Amoria undulata* Lamarck), the Courtly Cone (*Conus aulicus* Linné), and the the Textile Cone (*Conus ebraeusLinne).*

Recently, the observation of extremely long transients in lattice dynamical systems has brought into question the general relevance of attractors to complex spatio-temporal behavior.[37] Recalling our previous remarks concerning the use of transients, we conclude that even for these long-transient systems we may still obtain equations of motion and so make predictions of the complex transient behavior.

In closing this section, we suggest that the generalization of EoM analysis to highly interconnected systems, such as neural networks, autocatalytic networks, and massively parallel computers running particular algorithms, may provide a tool for investigating the informational architecture of these systems and their processing performance. This will hopefully allow for more quantitative understanding of these complex high-dimensional systems.

15. Prediction and control

Deducing the deterministic portion of a chaotic signal is one example of the general problem of detecting structure in data. The method's successes hint at an extension of nonlinear dynamics to pattern recognition, data compression, prediction, and control. From the perspective of dynamical systems theory, essentially the same problem occurs in these fields: there may be an apparent statistical component to a signal that contains some *a priori* structure of deterministic origin. This indicates that it may be reduced to a more compact, algorithmic specification. This specification forms its signature, which may then be incorporated into a prediction or control system. *

As discussed above, the estimated dynamic can be used to predict the next state $\vec{y}_n$, given an observed state $\vec{x}_n$ for discrete time behavior. The prediction is simply $\vec{F}(\vec{x}_n)$. $\vec{\sigma}_{obs}$ then measures the predictor's average effectiveness over the corresponding single sample time. For continuous time behavior Eq. (10.1) is a predictor for time T. $\vec{\sigma}_{obs}(T)$ measures the continuous time predictor's average effectiveness over time T. When $\vec{\sigma}_{obs}$ is small, the predictions are accurate; near unity, the predictions are no better than random guessing. Such *mean square error* measures of a predictor are standard tools in time series analysis. The estimated dynamic is a predictor that employs information from data points arbitrarily separated in the data series. This distinguishes it from typical "moving average" predictors.

Predictions for longer times are clearly possible. The information about the observed state, however, decays initially at a rate given by the metric entropy. This places a limit on the effectiveness of the observed state information for predictions over successively longer times. The maximum prediction time is simply $-\dfrac{\log_2 \epsilon}{h_\mu}$.[25, 39, 50]

*This section is an elaboration of a dynamical system predictor using dynamic estimation proposed[47] as an alternative to Pope's attempt[48] to predict iterations of the logistic map using Holland's classifier system.[49]

The possibility of automatically deducing equations of motion for general nonlinear dynamical systems suggests a new look at the design of prediction and control systems. EoM analysis can be used to design nonlinear prediction and control systems using data from a target system or some desired behavior to deduce the appropriate equations.† The estimated equations form an "internal" model. In this context, the model entropy $I(M)$ can be used as an informational measure of performance for such systems. It indicates how well-adapted the internal model is to the environment the system is attempting to predict or control.

A predictor, aside from monitoring the prediction error $\vec{\sigma}_{obs}$, could also adapt its internal model in ways to minimize $I(M)$, thereby improving its effectiveness over time. Accumulates errors in short-term prediction provide the required information for adaptation of the internal model over longer times. During the course of the adaptation, the system would simultaneously deduce nonlinear models while it used them to predict behavior.

Another more difficult, but illustrative, application of EoM-based prediction is to weather forecasting. During the inception of prediction theory, Wiener [51] discussed this from the point of view of modern ergodic theory and dynamics. Weather forecasting, as we envision it now, combines aspects of spatially-extended dynamics discussed above with EoM-based prediction. In setting up a weather forecaster, the initial internal model would derive from basic meteorology. The input would be historical and real-time data from temporal and spatial probes in satellites, earth-based weather stations, and other sources. Such a forecaster would require enormous computational and data acquisition resources. Nonetheless, with an appreciation of complex nonlinear dynamics built into it, there appears at present no *fundamental* physical or computational limitation to its operation.

One possible approach to the practical limitation of computational resources is to model the earth's weather system with a hierarchical and spatially-distributed set of EoM-based forecasters. Each level in the hierarchy would be assigned a particular spatial array of EoM forecasting subsystems, each one of which forecasts the weather in some localized spatial region for some specified range of spatial-wavenumber within that region. The main theoretical uncertainty in such a massively-parallel forecasting system would be the manner in which each local forecaster interpreted the predictions of nearby forecasters. One obvious implementation is to include all of the neighbors' variables. However, it may be necessary to include only those variables associated with the spatial transmission of information and not the entire state of each neighboring forecaster.

There are several benefits to a forecasting system of this design. First, it would adapt its internal model, improving its performance. Second, it would deduce over time the relevant observables and meteorological dynamics. Third, it could allocate computational resources where they were needed in areas of high information

† Reference [32] discusses this possibility at some length.

production. More local forecasters would be assigned to storm fronts and moved away from calm weather regions. A massively-parallel forecaster would be ideally suited to the new generation of parallel computers currently becoming available.

16. Scientific modeling and the dynamic of inquiry

Beyond simply producing an "experimental" model to be compared with theoretical models, the long-term goal of this endeavor is to make use of qualitative information of the type we have exploited to identify physically relevant variables and the new "laws" underlying their interaction. The hope is that a large fraction of the process of scientific investigation could be implemented automatically, without human intervention. Optimality criteria based on the model entropy $I(M)$ and its future derivatives will be central in this to select between competing theories.

Germane to this line of discussion, we recall Packard's algorithmic picture of scientific inquiry.* Traditionally, in a classical mechanical universe, there has been the tacit assumption of Baconian convergence of successively-refined models to those which predict *detailed* behavior, such as the future evolution of a system's state. Once a model predicts this detailed behavior it has been validated. When investigating nonlinear processes, one concludes that the existence of chaotic, deterministic behavior precludes the detailed comparison of theoretical models to experimental data. The conventional picture of inexorable improvement of models only applies to non-chaotic behavior.

Within a small sphere of scientific inquiry, such as deducing deterministic structure in a noisy data stream, EoM analysis allows for the complete modeling of Packard's scientific algorithm. The process of model improvement is given explicit form in terms of searching for optimal models in the space of dynamical system **D**. The model entropy provides a quantitative foundation for the discussion of the breakdown of the traditional scientific method for chaotic systems. EoM analysis suggests an alternative, and *convergent*, scientific algorithm for modeling time-dependent behavior based on the minimum model entropy criterion.

The selection between competing theories is based on minimizing the prediction error $\vec{\sigma}_{obs}$ and increasing the simplicity of the model. The model entropy is one combined measure of this. Kemeny[18] discusses a similar measure of a scientific theory's complexity that is based on the order of polynomials and how this must be traded off against a theory's prediction accuracy.

Our EoM model M can be thought of as a *scientific theory* in the sense that predictions can be based on it and their success evaluated. The evaluations in turn form the basis for comparison and validation. This might seem to elevate the rather humble notion of model that we have employed up to this point to an inappropriate generality. One should keep in mind, however, that we are discussing this in the limited context of

---

*Reference[7] discusses this in more detail. The modeling methodology of Box and Jenkins[52] should be compared with this picture.

noisy times series of combined stochastic and dynamic origin. As an example of the competition of scientific theories in a similar setting, consider the physical problem of fluid turbulence. A decade and a half ago there were the two rival theories due to Landau[53] and Lorenz-Ruelle-Takens.[1, 54] The relative weighing of these two hypothesized dynamic models has occurred since then, with the latter being the most appropriate in the onset regime of turbulent flows. We note now, at this late date, that this scientific evaluation could have been done automatically by deducing the equations of motion from the turbulence data and noting that the model was not of the Landau high-dimensional torus type, but was in fact of the low-dimensional chaotic attractor type. Even if the Landau model made equally accurate predictions, it is an inherently more complex model, $I(M_{Landau}) > I(M_{LRT})$, as it requires arbitrarily many oscillator subsystems.

We return to the choice between global and atlas equations of motion as it presents some interesting philosophical issues. As we have already indicated, in deducing equations of motion global EoM are to be preferred over atlas EoM, when the former is available. We have, nonetheless, demonstrated with the graph of the Duffing oscillator Poincaré map a dynamic that is very complex in any conventional global basis. Thus, atlas analysis may very well be our only recourse, since it adapts to unconventional nonlinearities. But an atlas EoM, with its set of charts and parameters for each, does not naturally indicate simplicity in the observed behavior, even if we incorporate efficient storage and evaluation. The size of the data structure is large, roughly $O(\epsilon^{-m_{bed}})$. We doubt that new "laws" could be discovered with this. If this turns out to be the typical situation as one applies EoM analysis to wider-ranging problems, then what of Poincaré's, and many other scientists', belief in nature's simplicity, or at least in humanity's ability to find simplicity in nature? According to Poincaré[55] it is by the identification of this simplicity that science progresses. Will we be left with the data of our experience coded into enormous, efficiently-organized data structures that admit no further simplification and from which humans cannot deduce order?

We conclude with a final query that encapsulates the problem discussed in this section. This concerns the automatic, machine-based deduction of macroscopic variables given only microscopic information about a thermodynamic system. Jaynes[56] addresses the complementary problem of prediction amongst macroscopic variables and the constraints imposed on microscopic states. Consider a box of gas for which we wish to deduce the existence and form of macroscopic variables, such as temperature, pressure, and entropy. The only available information, however, details the microscopic velocities and positions of each molecule. The theoretical questions then are, (i) what structures in the extremely high dimensional state space indicate macroscopic variables? And, can a machine (algorithm) identify these macroscopic variables automatically? Constructive, affirmative answers to these questions would be a major step toward automating scientific discovery, realizing Poincaré's "scientific machine".[55] Perhaps, however, the existence of such macroscopic structure can only be intuited.

17. Closing remarks

In a sense, we have reduced the problem of deducing equations of motion to statistical "quadrature". We have identified the necessary concepts from dynamical systems theory for the statistical problem of modeling data series. What is needed to go beyond the usefulness of the approach presented here is a theory of the relative complexity of functions. The rigorous foundation of EoM analysis requires a measure of the complexity of a function that can be applied to the graph of the dynamic represented by the data. The model entropy and the diagnostic are steps in this direction, but they are indirect and incomplete as they do not take into account the complexity of the chosen function basis itself.

In this essay we have been able to cover our topic only with broad strokes, from estimating experimental equations of motion to consideration of Poincaré's scientific machine. There are a number future problems that deserve closer scrutiny. The following lists a few.

(1) The relationship between the various selection criteria: minimum prediction error, minimum model size, maximum likelihood, maximum entropy, and minimum model entropy;

(2) Scaling of the model entropy with data set size and approximation type and order;

(3) The implementation of efficient hierarchical data structures for smooth atlases;

(4) The design of a generic EoM-based control systems;

(5) The design of the proposed massively-parallel forecaster; and finally,

(6) Adaptive noise reduction using EoM analysis.

With EoM analysis and the independent work in references [32] and [33] we see a rapprochement of dynamics, prediction, and modeling. Kolmogorov and Wiener[51] are generally credited with initiating prediction theory. It is a somewhat curious fact that while both were contributors to dynamical systems, especially Kolmogorov, the actual development of prediction theory has strayed quite far from the geometric state space approach of dynamics. We recall a similar lament[57] concerning chaotic dynamical systems, computation, and complexity theory. There too the parallel developments of dynamical systems and algorithmic complexity diverged rather far apart, and left physics altogether. During Kolmogorov's, Shannon's, and Wiener's day, dynamics, complexity, and physics were not disparate endeavors. It strikes us that the intimate connection between dynamics, one the one hand, and modeling, prediction, and complexity, on the other, has been ignored with no small intellectual cost. We have attempted, within a rather limited context, to aid in the simultaneous recurrence of these only-apparently distinct fields.

Acknowledgements

This research was funded in part by the ONR grant number N00014-86-K-0154. A post-doctoral fellowship from International Business Machines, Corporation, supported JPC in this work.

JPC thanks Doyne Farmer for discussing his independent work on prediction.

References

[1] E. N. Lorenz, "Deterministic Nonperiodic Flow", Journal of Atmospheric Sciences **20** (1963) 130.

[2] E. N. Lorenz, "The Problem of Deducing the Climate From the Governing Equations", Tellus **XVI** (1964) 1.

[3] J. P. Crutchfield, J. D. Farmer, N. H. Packard, and R. S. Shaw, "Chaos", Scientific American **255** (December 1986) 46.

[4] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a Time Series", Physical Review Letters **45** (1980) 712.

[5] G. Mayer-Kress, *Dimensions and Entropies in Chaotic Systems,* (Springer-Verlag, Berlin, 1986).

[6] F. Takens, "Detecting Strange Attractors in Fluid Turbulence", Lecture Notes in Mathematics **898** (1981) 366.

[7] J. P. Crutchfield, *Noisy Chaos,* (Dissertation, University of California, Santa Cruz, 1983).

[8] M. Aoki, *State Space Modeling of Time Series,* (Springer-Verlag, Berlin, 1987).

[9] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov Exponents from a Time Series", Physica **16D** (1985) 285.

[10] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, "Liapunov Exponents from Time Series", IHES preprint **P/86/20** (April 1986).

[11] M. Sano and Y. Sawada, "Measurement of the Lyapunov Spectrum from Chaotic Time Series", Physical Review Letters **55** (1985) 1082.

[12] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication,* (University of Illinois Press, Champaign-Urbana, 1962).

[13] D. S. Broomhead and G. P. King, "Extracting Qualitative Dynamics from Experimental Data", Physica **20D** (1986) 217.

[14] H. Froehling, J. P. Crutchfield, J. D. Farmer, N. H. Packard, and R.S. Shaw, "On Determining the Dimension of Chaotic Flows", Physica **3D,** (1981) 605.

[15] J. P. Crutchfield, "Noise, Scaling, Entropy, and Dimension", in preparation (1986).

[16] H. Whitney, "The Self-Intersections of a Smooth N-Manifold in 2N-Space", Annals of Mathematics **45** (1944) 220.

[17] V. Guilleman and A. Pollack, *Differential Topology,* (Prentice-Hall, Englewood Cliffs, New Jersey, 1974).

[18] J. G. Kemeny, "The Use of Simplicity in Induction", Philosophical Review **62** (1953) 391.

[19] E. T. Jaynes, “Where do We Stand on Maximum Entropy?” in *The Maximum Entropy Formalism*, edited by R. D. Levine and M. Tribus, (MIT Press, Cambridge, 1978).

[20] H. Akaike, “A New Look at the Statistical Model Identification”, IEEE Transactions on Automatic Control **AC-19** (1974) 716.

[21] J. P. Crutchfield, *Prediction and Stability in Classical Mechanics,* (Senior Thesis, University of California, Santa Cruz, 1979).

[22] I. Shimada and T. Nagashima, “A Numerical Approach to Ergodic Problem of Dissipative Dynamical Systems”, Progress of Theoretical Physics **61** (1979) 1605.

[23] J. D. Farmer, E. Ott, and J. A. Yorke, “The Dimension of Chaotic Attractors”, Physica **7D** (1983) 153.

[24] Y. Oono, “A Heuristic Approach to the Kolmogorov Entropy as a Disorder Parameter”, Progress of Theoretical Physics **60** (1978) 1944.

[25] J. P. Crutchfield and N. H. Packard, “Symbolic Dynamics of Noisy Chaos”, Physica **7D** (1983) 201.

[26] D. Ruelle, “Small Random Perturbations of Dynamical Systems and the Definition of Attractors”, Communications of Mathematical Physics **82** (1981) 137-151.

[27] J. H. Wilkinson and C. Reinsch, *Linear Algebra, Handbook for Automatic Computation, vol. II,* (Springer-Verlag, Berlin, 1971).

[28] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes, The Art of Scientific Programming,* (Cambridge University Press, Cambridge, 1986).

[29] D. R. J. Chillingworth, *Differential Topology with a View to Applications,* (Pitman, London, 1976).

[30] K. Fukunaga and D. R. Olsen, “An Algorithm for Finding Intrinsic Dimensionality of Data”, IEEE Transactions on Computers **C-20** (1971) 176.

[31] M. B. Priestley, “State-Dependent Models: A General Approach to Non-linear Time Series Analysis”, Journal of Time Series Analysis **1** (1980) 47.

[32] S. Omohundro, “Efficient Algorithms with Neural Network Behavior”, Journal of Complex Systems **1** (1987) 273.

[33] J. D. Farmer and J. Sidorowich, “Predicting Chaotic Time Series”, LANL preprint LA-UR-87-1502 (1987).

[34] J.-P. Eckmann and D. Ruelle, “Ergodic Theory of Chaos and Strange Attractors”, Reviews of Modern Physics **57** (1985) 617.

[35] O. E. Rossler, “An Equation for Hyperchaos”, Physics Letters **71A** (1979) 155.

[36] M. Henon and C. Heiles, “The Applicability of the Third Integral of Motion: Some Numerical Experiments”, Astronomical Journal **69** (1964) 73.

[37] J. P. Crutchfield and K. Kaneko, “Phenomenology of Spatio-Temporal Chaos” in *Directions in Chaos*, edited by Hao Bai-lin, (World Scientific Publishing Co.,

Singapore, 1987).

[38] J. P. Crutchfield, unpublished (1979). These results were posted and on display from that time until 1983 at the Physics Department, University of California, Santa Cruz.

[39] R. Shaw, “Strange Attractors, Chaotic Behavior, and Information Flow”, Zeitschrift fur Naturforshung **36a** (1981) 80.

[40] J. P. Crutchfield, *Chaotic Attractors of Driven Oscillators (16mm film),* (Aerial Press, Box 1360, Santa Cruz, CA 95061, 1982).

[41] P. Martien, S. C. Pope, P. L. Scott, and R. S. Shaw, “The Chaotic Behavior of the Leaky Faucet”, Physics Letters **110A** (1985) 399.

[42] J.P. Crutchfield, J.D. Farmer, and B.A. Huberman, “Fluctuations and Simple Chaotic Dynamics”, Physics Reports **92** (1982) 45.

[43] J. P. Crutchfield and N. H. Packard, “Observing Spatial Dynamics”, presented at the International Workshop on Dimensions and Entropies in Chaotic Systems (11-16 September 1985) Pecos, New Mexico.

[44] J. P. Crutchfield and K. Kaneko, “Space-Time Information Theory”, in preparation .

[45] S. Wolfram, “Universality and Complexity in Cellular Automata”, Physica **10D** (1984) 1.

[46] J. P. Crutchfield and K. Kaneko, “Statistical and Deterministic Mechanics of Spatio-Temporal Chaos”, in preparation .

[47] J. P. Crutchfield, “Quantifying Chaos with Predictors”, El Rancho Institute (1985) unpublished notes.

[48] S. Pope, El Rancho Institute (1985) private communication.

[49] J. H. Holland, “A Mathematical Framework for Studying Learning in Classifier Systems”, Physica **22D** (1986) 307.

[50] J. D. Farmer, J. P. Crutchfield, H. Froehling, N. H. Packard, and R.S. Shaw, “Power Spectra and Mixing Properties of Strange Attractors”, Annals of the New York Academy of Sciences **357** (1980) 453.

[51] N. Wiener, “Nonlinear Prediction and Dynamics” in *Norbert Wiener, Collected Works III*, (MIT Press, Cambridge, 1981).

[52] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control,* (Holden-Day, San Francisco, 1970).

[53] L. D. Landau and E. M. Lifshitz, *Fluid Dynamics,* (Pergamon Press, London, 1959).

[54] D. Ruelle and F. Takens, “On the Nature of Turbulence”, Communications of Mathematical Physics **20** (1971) 167.

[55] H. Poincare, *Science and Hypothesis,* (Dover, New York, 1952).

[56] E. T. Jaynes, “Macroscopic Prediction” in *Complex Systems - Operational Approaches*, edited by H. Haken, (Springer-Verlag, Berlin, 1985) 254.

[57] J. P. Crutchfield and N. H. Packard, "Symbolic Dynamics of One-Dimensional Maps: Entropies, Finite Precision, and Noise", International Journal of Theoretical Physics **21** (1982) 433.

Table of Contents